Development of a Machine Learning Classifier to retrieve Time-Series of Ash Componentry at Tungurahua Volcano, Ecuador (1999-2016)

Gabriel Adler Cancino[1,2], Damià Benet[3], Chiara Maria Petrone[1], H. Elizabeth Gaunt[2], Alexander L Steele[2], Benjamin Bernard[4]

gabriel.adlercancino@nhm.ac.uk

1. Natural History Museum, Volcano Petrology Group, Cromwell Road, London, UK, SW7 5BD
2. Department of Earth Sciences, University College London, Gower Street, London, UK, WC1E 6BT
3. Institut de Physique du Globe de Paris, Université Paris Cité, Paris, France
4. Instituto Geofísico, Escuela Politécnica Nacional, Quito, Ecuador

1. What is componentry?

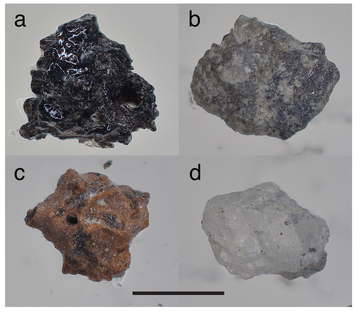


Fig 1: a - Juvenile grain (formed in fresh magma, glassy lustre, vesicular, angular); b - Lithic grain (entrained from conduit wall, dull lustre, sub-rounded, no glass); c - Altered grain (oxidised or hydrothermally altered, red, brown or white coatings); d - Free crystal (origin unknown, faceted, white plagioclases or dark green pyroxenes)

- The classification of ash grains into 'components'.
- Classification schemes vary a lot. Figure 1 shows a commonly used 4-component scheme.
- The presence of juvenile or altered grains imply the presence of fresh magma or hydrothermal contact, respectively, into the system, helping observatories to pre-empt eruptive style.
- Manual classification can be time-consuming and subjective, so Machine Learning (ML) can make this technique more viable for near-real-time monitoring.

2. Geological Context and Sampling

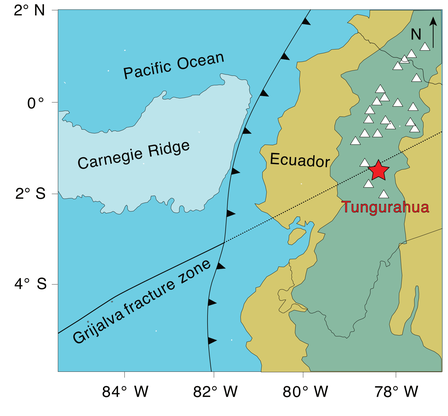
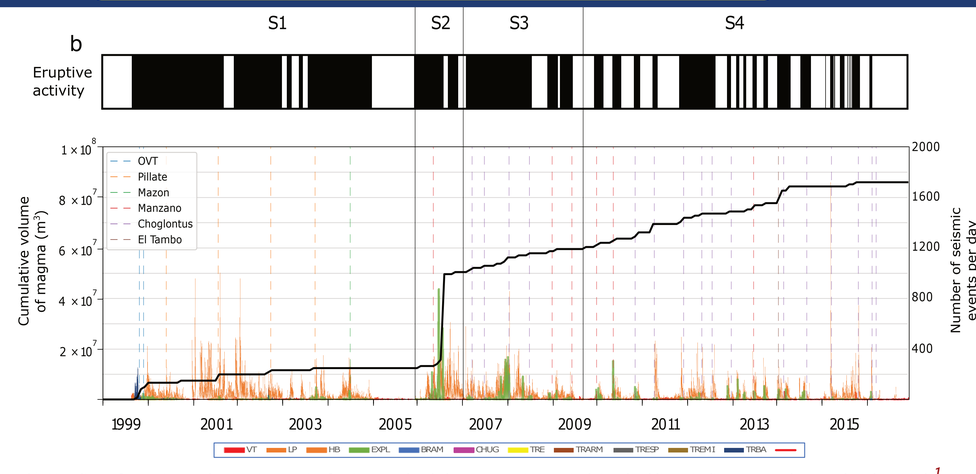


Fig 2: a) Map of the Northern Andean Volcanic Zone[1]



b) Timeline of eruptive and seismic activity for Tungurahua's latest eruptive period. Dashed lines indicate samples used in this study.[1]

- Tungurahua is part of the Northern Andean Volcanic Zone.
- In its modern period (3000BP - Present), Tungurahua has erupted ~1/100 years. The latest eruption lasted from 1999-2016.
- The eruption saw a mix of Strombolian, Vulcanian and Sub-Plinian events.
- Ash samples were collected concurrently with geophysical monitoring data, making Tungurahua a great case study for an ash time-series for petrological monitoring.

3. In-depth Grain Identification

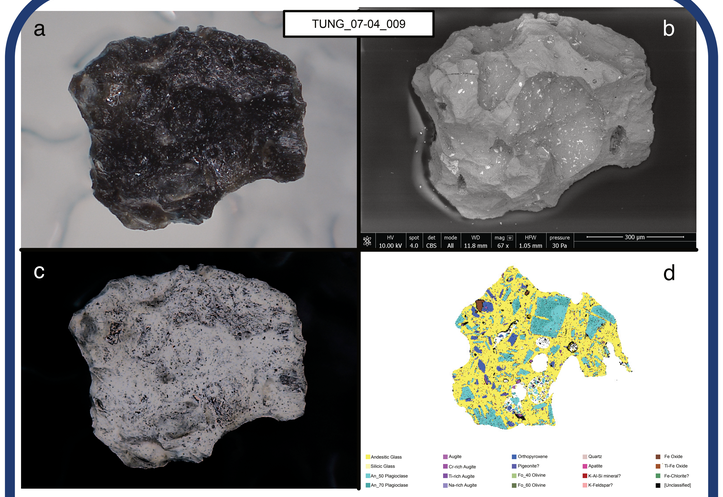


Fig 3: a - Optical microscope image (HIROX HRX-01); b - Scanning Electron Microscope surface image (FEI Quanta 650 FEG); c - Image a with inverted colours (easier for classification); d - Phase map of a cross-section of the grain (TESCAN TIMA)

- 160 grains from 16 samples were imaged by 3 different microscopes to understand the petrogenesis of different grain types and to identify components 'objectively'. At least 8 grain types were identified for Tungurahua.
- This technique is too detailed to be used for regular monitoring, but has been used to test further ML techniques.

4. Unsupervised Learning - t-SNE

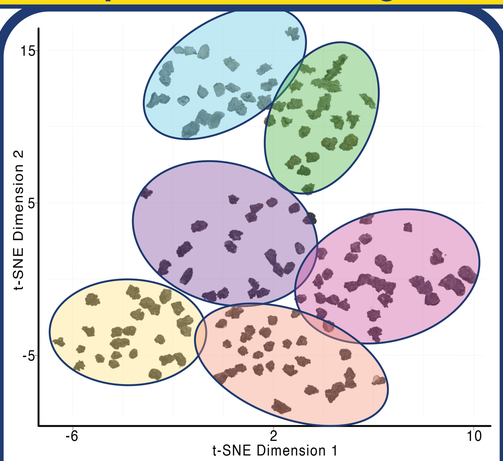


Fig 4: t-SNE visualisation of grains. Coloured clusters identified by k-means clustering

- Grain images were segmented, and 40 shape, textural and colour features were extracted. t-SNE dimensionality reduction was used to visualise the distribution of features across grains.
- Grains cluster into components relatively well. Feature engineering will be needed to improve accuracy.

5. Supervised Learning - VolcAshDB

Fig 5: VolcAshDB classification example for sample TUNG_07-04. Bars show prediction likelihood for each component. Red circles show that the sample set is not similar to the training set.

- Classification was done by the VolcAshDB[2] open-source model under the 4-component scheme in Figure 1.
- The 'out-of-the box' model over-predicted false positives for lithic and free crystal components. Training on Tungurahua samples will be needed to improve accuracy.

6. Synthesis

Unsupervised

- Quick and easy-to-use. Grains can be imaged and processed in 24 hours. Applicable for monitoring.
- No need for a pre-determined classification scheme, so can better visualise the full range of grain types in the eruption.
- Clusters are not comparable for every sample, so proportions cannot be quantified for time-series.

Supervised

- Very quick and easy-to-use once trained. The extent of training needed may depend on the samples and the observatory's facilities.
- Quantifiable and classifies with higher accuracy than unsupervised learning once trained.
- Results are subject to engrained biases in the classification scheme and training dataset.

1. Vlastélic, I., Sainlot, N., Samaniego, P., Bernard, B., Nauret, F., Hidalgo, S., Auclair, D. and Gannoun, A. (2023). 'Arc volcano activity driven by small-scale metasomatism of the magma source'. Nature Geoscience. Nature Publishing Group, 16 (4), pp. 363–370.
2. Benet, D., Costa, F. and Widiwijayanti, C. (2024). 'Volcanic Ash Classification Through Machine Learning'. Geochemistry, Geophysics, Geosystems, 25 (3).

Natural Environment Research Council
Ciencia y Tecnología
Secretaría de Ciencia, Humanidades, Tecnología e Innovación